

DR. ALEXANDER M. FRASER - Curriculum Vitae

Universität Stuttgart - School of Computer Science, Electrical Engineering and Information Technology
Institute for Natural Language Processing +49 711 685-81375 (phone)
Statistical Natural Language Processing Group fraser@ims.uni-stuttgart.de
Azenbergstrasse 12, 70174 Stuttgart, Germany www.ims.uni-stuttgart.de/~fraser

Education

University of Southern California Los Angeles CA, USA
PhD in Computer Science defended 07/2007
– Thesis Title: Improved Word Alignments for Statistical Machine Translation
– Advisor: Daniel Marcu
– Committee: Eduard Hovy, Gareth James, Kevin Knight, Daniel Marcu, Paul Rosenbloom
– MSc (2002) in Computer Science
– GPA 3.86/4.00

University of Massachusetts/Boston Boston MA, USA
BA in Computer Science and Economics, Concentration in Linguistics 05/1995
– GPA 3.72/4.00

Research Interests

Natural Language Processing: statistical machine translation, hybrid machine translation, statistical parsing of syntax, morphology, cross-language information retrieval, language modeling, automatic evaluation, discourse

Machine Learning: structured prediction, semi-supervised learning, unsupervised learning, hybrid discriminative and generative models, distributed algorithms for machine learning

Professional Experience

Institute for Natural Language Processing (IfNLP/IMS) Stuttgart, Germany
University of Stuttgart
Senior Researcher, Leader of Statistical Machine Translation Group 10/2009 to present
DFG Project: Models of Morphosyntax for Statistical Machine Translation

Postdoctoral Researcher 11/2007 to 09/2009
DFG Collaborative Research Centre 732: Incremental Specification in Context

Information Sciences Institute (ISI) Marina del Rey CA, USA
University of Southern California
Graduate Research Assistant 09/2000 to 09/2007
(except as indicated below)

Language Weaver Marina del Rey CA, USA
Research Scientist 06/2003 to 05/2004

Department of Speech and Language Processing Cambridge MA, USA
BBN Technologies
Research Scientist 06/2002 to 08/2002
05/2001 to 08/2001
02/2000 to 08/2000

Satellife Boston MA, USA
Technical Director 12/1995 to 01/1998
Systems Programmer 09/1995 to 12/1995
UNIX System Administrator 03/1994 to 12/1994

DR. ALEXANDER M. FRASER - Curriculum Vitae

The Forum Corporation

Programmer/Analyst

Programmer

Boston MA, USA

02/1988 to 01/1990

05/1987 to 01/1988

Publications

Refereed Journal Publications

- **Alexander Fraser**, Daniel Marcu (2007). Measuring Word Alignment Quality for Statistical Machine Translation. *Computational Linguistics*, 33(3): 293-303

Book Chapters

- **Alexander Fraser**, William Wong (2008). The Language Weaver Statistical Machine Translation Software System. In A. Farghaly, editor, *Arabic Computational Linguistics*, CSLI Publications, Chapter 9, pages 257 to 288.

Refereed Conference Publications

- **Alexander Fraser**, Marion Weller, Aoife Cahill, Fabienne Cap (2012). Modeling Inflection and Word-Formation in SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, to appear.
- Anita Gojun, **Alexander Fraser** (2012). Determining the Placement of German Verbs in English-to-German SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, to appear.
- Hassan Sajjad, Nadir Durrani, Helmut Schmid, **Alexander Fraser** (2011). Comparing Two Techniques for Learning Transliteration Models Using a Parallel Corpus. In *Proceedings of The 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 129 to 137, Chiang Mai, Thailand, November.
- Nadir Durrani, Helmut Schmid, **Alexander Fraser** (2011). A Joint Sequence Translation Model with Integrated Reordering. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1045 to 1054, Portland, Oregon, USA, June.
- Hassan Sajjad, **Alexander Fraser**, Helmut Schmid (2011). An Algorithm for Unsupervised Transliteration Mining with an Application to Word Alignment. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 430 to 439, Portland, Oregon, USA, June.
- Fabienne Braune, **Alexander Fraser** (2010). Improved Unsupervised Sentence Alignment for Parallel and Asymmetrical Bilingual Corpora. In *Proceedings of the the 23rd International Conference on Computational Linguistics (COLING) - Posters*, pages 81 to 89, Beijing, China, August.
- Nadir Durrani, Hassan Sajjad, **Alexander Fraser**, Helmut Schmid (2010). Hindi-to-Urdu Machine Translation Through Transliteration. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 465 to 474, Uppsala, Sweden, July.
- Florian Schwarck, **Alexander Fraser**, Hinrich Schütze (2010). Bitext-Based Resolution of German Subject-Object Ambiguities. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT) - Short Paper, Oral Presentation*, pages 737 to 740, Los Angeles, California, USA, June.

DR. ALEXANDER M. FRASER - Curriculum Vitae

- **Alexander Fraser**, Renjing Wang, Hinrich Schütze (2009). Rich Bitext Projection Features for Parse Reranking. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 282 to 290, Athens, Greece, March.
- Hamidreza Kobdani, **Alexander Fraser**, Hinrich Schütze (2009). Word Alignment by Thresholded Two-Dimensional Normalization. In *Proceedings of MT Summit XII - Posters*, pages 260 to 267, Ottawa, Canada, August.
- **Alexander Fraser**, Daniel Marcu (2007). Getting the Structure Right for Word Alignment: LEAF. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 51 to 60, Prague, Czech Republic, June.
- **Alexander Fraser**, Daniel Marcu (2006). Semi-Supervised Training for Statistical Word Alignment. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pages 769 to 776, Sydney, Australia, July.
- Dragos Stefan Munteanu, **Alexander Fraser**, Daniel Marcu (2004). Improved Machine Translation Performance via Parallel Sentence Extraction from Comparable Corpora. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL)*, pages 265 to 272, Boston, Massachusetts, USA, May.
- Franz Josef Och, Daniel Gildea, Sanjeev Khudanpur, Anoop Sarkar, Kenji Yamada, **Alex Fraser**, Shankar Kumar, Libin Shen, David Smith, Katherine Eng, Viren Jain, Zhen Jin, Dragomir Radev (2004). A Smorgasbord of Features for Statistical Machine Translation. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL)*, pages 161 to 168, Boston, Massachusetts, USA, May.
- Jinxi Xu, **Alexander Fraser**, Ralph Weischedel (2002). Empirical Studies in Strategies for Arabic Retrieval. In *Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 269 to 274, Tampere, Finland, August.

Unrefereed Conference Publications

- **Alexander Fraser**, Renjing Wang, Hinrich Schütze (2009). A progress report on bitext parsing. In *2nd International Conference on the Future of Information Sciences (InFuture): Digital Resources and Knowledge Sharing*, pages 43 to 52, Zagreb, Croatia, November. Invited Paper.

Refereed Workshop Publications

- Fabienne Fritzing, **Alexander Fraser** (2010). How to Avoid Burning Ducks: Combining Linguistic Analysis and Corpus Statistics for German Compound Processing. In *Proceedings of the ACL 2010 Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 224 to 234, Uppsala, Sweden, July.

Refereed Workshop Presentations

- **Alexander Fraser**, Marion Weller, Aoife Cahill, Fabienne Cap (2011). Morphological Generation of German for SMT. *Abstract accepted for presentation at Machine Translation and Morphologically-rich Languages. Research Workshop of the Israel Science Foundation*. University of Haifa, Israel, January, 2011.

DR. ALEXANDER M. FRASER - Curriculum Vitae

Shared Task Workshop Publications

- **Alexander Fraser** (2009). Experiments in Morphosyntactic Processing for Translating to and from German. In *Proceedings of the EACL 2009 Fourth Workshop on Statistical Machine Translation*, pages 115 to 119, Athens, Greece, March.
- **Alexander Fraser**, Daniel Marcu (2005). ISI's Participation in the Romanian-English Alignment Task. In *ACL 2005 Workshop on Building and Using Parallel Texts*, pages 91 to 94, Ann Arbor, Michigan, USA, June.
- **Alexander Fraser**, Jinxi Xu, Ralph Weischedel (2002). TREC 2002 Cross-lingual Retrieval at BBN. In *Voorhees, E.M. & Harman, D.K. (Eds.) The Eleventh Text Retrieval Conference (TREC)*, pages 102 to 106, NIST Special Publication 500-251.
- Jinxi Xu, **Alexander Fraser**, Ralph Weischedel (2001). TREC 2001 Cross-lingual Retrieval at BBN. In *Voorhees, E.M. & Harman, D.K. (Eds.) The Tenth Text Retrieval Conference (TREC)*, pages 68 to 77, NIST Special Publication 500-250.

Technical Reports

- **Alexander Fraser**, Daniel Marcu (2006). Measuring Word Alignment Quality for Statistical Machine Translation. Technical Report ISI-TR-616, ISI/University of Southern California, May.
- Franz Josef Och, Daniel Gildea, Sanjeev Khudanpur, Anoop Sarkar, Kenji Yamada, **Alex Fraser**, Shankar Kumar, Libin Shen, David Smith, Katherine Eng, Viren Jain, Zhen Jin, Dragomir Radev (2003). Syntax for Statistical Machine Translation. Johns Hopkins University 2003 Summer Workshop on Language Engineering, Center for Language and Speech Processing, Baltimore, Maryland, USA, August.

Patent Application

- **Alexander Fraser**, Daniel Marcu. Semi-Supervised Training for Statistical Word Alignment. Application 11/592,450 United States Patent and Trademark Office, filed 11/2006

Honors / Awards

- University of Southern California - Charles Lee Powell Graduate Fellowship, full support and additional stipend for 3 years of PhD studies
- Full sponsorship to spend 8 weeks in Baltimore at the Johns Hopkins University CLSP Summer Workshop 2003: Syntax for Statistical Machine Translation, led by Franz Och and Daniel Gildea
- 1000 Euro award for open source, distributed implementation of LEAF alignment model (from dissertation) at Euromatrix Second Machine Translation Marathon
- Full sponsorship for participation in Euromatrix Second Machine Translation Marathon (one week)
- SIGIR 2002 Student Travel Award
- University of Massachusetts - Graduated magna cum laude
- University of Massachusetts - Juan Carlos Merlo Prize, awarded annually to a single graduating mathematics or computer science student for outstanding achievement
- University of Massachusetts - Honors in Computer Science
- University of Massachusetts - Honors in Economics

Grants

- Principal Investigator on DFG 2-person, 3-year project, “Models of Morphosyntax for Statistical Machine Translation”
- Deputy Investigator, European Commission FP7 STREP, “Terminology Extraction, Translation Tools and Comparable Corpora”, Coordinator Beatrice Daille, Nantes
- Principal Investigator on grant for design, implementation and quality assurance of Language Weaver Arabic/English hybrid statistical machine translation system
- Principal Investigator on Small Business Innovation Research grant to Language Weaver, “Linguistic Error Analysis of Statistical Machine Translation”

Teaching

- “Statistical Machine Translation”, with Andreas Maletti, Seminar (Hauptseminar), University of Stuttgart, Winter Semester 2011/2012
- “Machine Learning for Natural Language Processing” Seminar (Hauptseminar), with Stefan Riezler and Hinrich Schuetze, Joint Seminar: University of Heidelberg and University of Stuttgart, Winter Semester 2010/2011
- “Statistical Machine Translation”, with Helmut Schmid, Seminar (Lektürekurs), University of Stuttgart, Summer Semester 2010
- “Parsing and Machine Translation”, with Helmut Schmid, Seminar (Hauptseminar), University of Stuttgart, Summer Semester 2009
- “Data-driven and Hybrid Approaches to Machine Translation”, with Martin Forst, intensive two week course at German Society for Linguistics Fall School in Konstanz, 2009
- Taught one week intensive course on Statistical Machine Translation at 9th Summer School of the European Masters in Language and Speech Technology Programme, 2008

Advising

Ph.D. Advising

- Fabienne Braune, 10/2009-
- Fabienne Cap (nee Fritzingler), 10/2009-
- Nadir Durrani, 04/2008-
- Hassan Sajjad, 04/2008-

Masters/Bachelors Advising

- Fabienne Braune (U. Freiburg, Switzerland), Masters in Computer Science, Analysing and Improving Unsupervised Sentence Alignment, completed 09/2009
- Fabienne Fritzingler, Diplom in Computational Linguistics, Extracting Multiword Expressions from Parallel Text, completed 08/2008
- Anita Gojun, Diplom in Computational Linguistics, Null Subjects in Statistical Machine Translation: A Case Study on Aligning English and Italian Verb Phrases with Pronominal Subjects, completed 08/2010

DR. ALEXANDER M. FRASER - Curriculum Vitae

- Ronny Jauch, Studienarbeit in Computational Linguistics, Umordnung von quellsprachlichen Sätzen in die Satzstruktur der Zielsprache (Reordering of Source-Language Sentences into the Ordering of the Target Language), completed 08/2011
- Renjing Wang, Diplom in Computer Science, Self-Training for German Parsing, completed 09/2011
- Renjing Wang, Studienarbeit in Computer Science, Parsing Bitext, completed 05/2009
- Marion Weller, Diplom in Computational Linguistics, An Empirical Analysis of Source Context Features for Phrase-Based Statistical Machine Translation, completed 08/2010
- Marion Weller, Studienarbeit in Computational Linguistics, Separate Morphologiebehandlung als Methode zur Verbesserung statistischer maschineller Übersetzung (Separate Morphological Processing as a Method for Improving Statistical Machine Translation), completed 05/2009. Winner of **Informatik-Verbund-Stuttgart Award** (500 Euro prize awarded twice each year for the best Studienarbeit in computer-related fields in Stuttgart)

External

- Examiner, “Licentiate of Philosophy” (similar to American Ph.D. qualification exam): Sara Stymne, Department of Computer and Information Science, Linköping University, Sweden. December 2009
- Erasmus Host: Marija Brkic (research on Discriminative Modeling for Machine Translation), Visiting Stuttgart from the Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia. September 2010 to February 2011

Invited Presentations

- Panel, “Arabic Natural Language Processing Technologies Round-Table”, Qatar Computing Research Institute, Doha, Qatar, 2011
- Invited Talk, “Analysis and Generation of German for Statistical Machine Translation”, Institute for Language, Cognition and Computation Seminar, University of Edinburgh, Scotland, 2010
- Keynote Speaker, “Morphosyntactic Correspondences and a Progress Report on Bitext Parsing”, InFuture 2009, Zagreb, Croatia, 2009
- Invited Talk, “Beyond Phrase-Based Statistical Machine Translation”, Gesellschaft für Sprachtechnologie und Computerlinguistik Workshop “Costs and Uses of Machine Translation”, 2009
- Invited Talk, “Semi-Supervised Word Alignment”, Research Track of Euromatrix Second Machine Translation Marathon, 2008
- Invited Talk, “Statistical Word Alignment”, Department of Computational Linguistics, University of Heidelberg, 2008
- Invited Talk, “Getting the Structure Right for Word Alignment: LEAF”, Linguistics Department, University of Potsdam, 2007
- Invited Talk, “Improved Word Alignments for Statistical Machine Translation”, Language Technology Laboratory, DFKI (German Research Center for Artificial Intelligence), 2006
- Invited Talk, “Improved Word Alignments for Statistical Machine Translation”, Institute for Natural Language Processing (IMS), University of Stuttgart, 2006
- Demonstration, Language Weaver Arabic/English Statistical Machine Translation System, HLT-NAACL 2004

DR. ALEXANDER M. FRASER - Curriculum Vitae

- Demonstration, Language Weaver Arabic/English Statistical Machine Translation System, COLING Workshop on Computational Approaches to Arabic Script-based Languages 2004
- Panel, “Issues in Arabic to English Machine Translation”, MT Summit IX Workshop on Machine Translation for Semitic Languages, 2003
- Invited Talk, “USC/ISI’s Arabic/English Machine Translation System”, US National Institute of Standards Machine Translation Evaluation Workshop, 2003

Professional Activities

Journal Reviewer

- Computational Linguistics
- Journal of Artificial Intelligence Research
- Artificial Intelligence

Conference/Workshop Reviewer

- COLING-ACL 2006, ACL-HLT 2008, ACL 2009, ACL 2010, ACL 2011
- NAACL 2004, NAACL-HLT 2007, NAACL-HLT 2010
- EACL 2009, 2012
- NIPS 2010, 2011
- EMNLP 2006, EMNLP-CoNLL 2007, EMNLP 2008, EMNLP 2009, EMNLP 2010, EMNLP 2011
- COLING 2008, COLING 2010
- MT Summit 2009
- IJCNLP 2008
- ACL/NAACL/EACL Workshops on Machine Translation: 2006, 2007, 2008, 2009, 2010, 2011
- ACL Workshop on Building and Using Parallel Texts: 2005
- EACL/ACL Workshops on Computational Approaches to Semitic Languages: 2005, 2007, 2009
- IWSLT International Workshop on Spoken Language Translation: 2005
- COLING Workshop on Human Judgements in Computational Linguistics: 2008
- TMI Theoretical and Methodological Issues in Machine Translation: 2004
- MT Summit IX Workshop on Machine Translation for Semitic Languages: 2003

Other Reviewing

- U.S. National Science Foundation - Proposal Reviewer
- Neural Information Processing Series, “Learning Machine Translation”, MIT Press
- Grace Hopper Celebration of Women in Computing, Anita Borg Institute

DR. ALEXANDER M. FRASER - Curriculum Vitae

Other Activities

- Designed and managed creation of the first commercially available statistical machine translation system (Language Weaver Arabic/English)
- Engineered top ranked systems in cross-language information retrieval, information retrieval, word alignment and machine translation
 - Best Arabic information retrieval, TREC 2001
 - Best English to Arabic cross-language retrieval, TREC 2001, TREC 2002
 - Best Arabic to English machine translation, NIST-MT 2003 constrained and unconstrained tracks
 - Best Romanian/English word alignment, ACL-WPT 2005 constrained track
 - Word alignments used in 1st or 2nd best ranked Chinese to English and Arabic to English machine translation systems, NIST-MT 2006, 2008, 2009
 - Best German to English machine translation, ACL-WMT 2009 constrained track
- Software provided to the research community:
 - Cube pruning algorithm implemented in open source Moses statistical machine translation engine, with team at Open Source track of Euromatrix Second Machine Translation Marathon
 - Word alignment graphical user interface implemented in Java Swing, used by 5 research labs
- Data provided to the research community:
 - Probabilistic Dictionary (with Jinxi Xu) for use in TREC-2002 Cross-Lingual Information Retrieval Task (using English queries to request Arabic documents), see <http://www.glue.umd.edu/~dlrg/clir/arabic.html>
 - BBN release of the Arabic/English UN corpus (with Jinxi Xu, Ralph Weischedel), distributed by the Linguistic Data Consortium
 - Sentence aligned the Ummah corpus, distributed by the Linguistic Data Consortium
- Organizing of Stuttgart Reading Group on Parsing and Machine Translation

General Skills / Qualifications

- Twelve years of experience in research positions
- Nine additional years of experience in information technology positions
- Programming in C++, Perl, Java, C, Python, LISP, SmallTalk. Implementation and optimization of distributed and multithreaded algorithms
- Experienced work group manager
- Skilled trainer of both technical and non-technical staff
- Native speaker of English. Fluent in French, German, Spanish. Proficient in Arabic

References

Available upon request